

Class_19

Hayoung A15531571

11/30/2021

Section 4 Homework

Let's load the file into our worksheet

```
expr <- read.csv("updated.csv")
```

```
head(expr)
```

```
##   X sample geno      exp
## 1 1 HG00367  A/G 28.96038
## 2 2 NA20768  A/G 20.24449
## 3 3 HG00361  A/A 31.32628
## 4 4 HG00135  A/A 34.11169
## 5 5 NA18870  G/G 18.25141
## 6 6 NA11993  A/A 32.89721
```

Q13. Read this file into R and determine the sample size for each genotype and their corresponding median expression levels for each of these genotypes.

```
table(expr$geno)
```

```
##
## A/A A/G G/G
## 108 233 121
```

There are 108 A|A, 233 A|G, and 121 G|G genotype samples.

```
library(plyr)
```

```
p_meds <- ddply(expr, .(geno), summarise, med = median(exp) )
p_meds
```

```
##   geno      med
## 1  A/A 31.24847
## 2  A/G 25.06486
## 3  G/G 20.07363
```

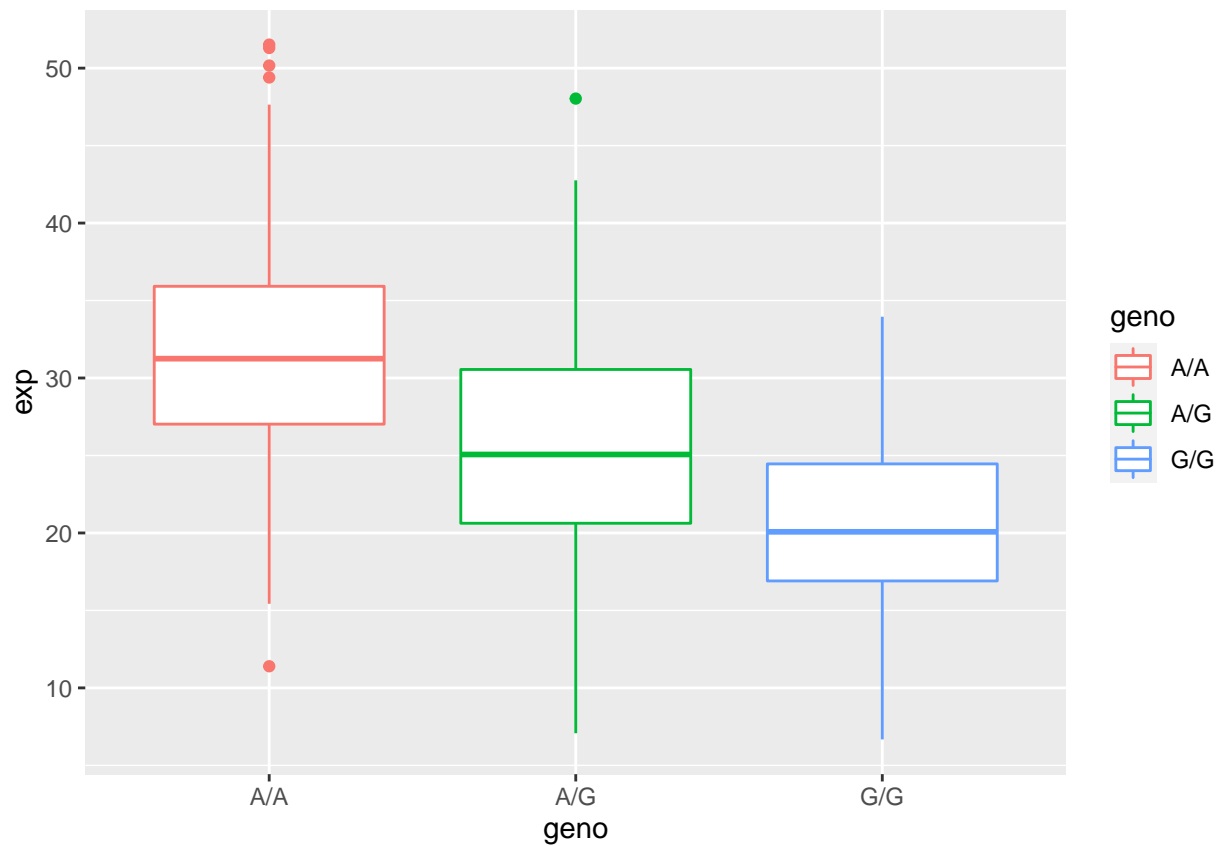
The median expressions of the genotypes are 31.25 for A|A, 25.06 for A|G, and 20.07 for G|G

Q14. Generate a boxplot with a box per genotype, what can you infer from the relative expression value between A|A and G|G displayed in this plot? Does the SNP effect the expression of ORMDL3?

```
library(ggplot2)

plot <- ggplot(expr) +
  aes(x = geno, y = exp, col = geno) +
  geom_boxplot()

plot
```



From the above plot, we can see that for A|A and G|G, the “boxes” do not overlap, and thus we can infer that they are relatively separate from one another, and there is a difference in expression. The SNP does effect the expression of ORMDL3, we can see a difference in expression when comparing genotypes.